ENHANCING SEISMIC CALIBRATION RESEARCH THROUGH SOFTWARE AUTOMATION AND SCIENTIFIC INFORMATION MANAGEMENT

Stanley Ruppert, Douglas A. Dodge, Michael D. Ganzberger, Teresa F. Hauk, and Eric M. Matzel

Lawrence Livermore National Laboratory

Sponsored by National Nuclear Security Administration

Contract No. DE-AC52-07NA27344

ABSTRACT

The National Nuclear Security Administration (NNSA) Ground-Based Nuclear Explosion Monitoring Research and Development (GNEMRD) Program at LLNL continues to make significant progress enhancing the process of deriving seismic calibrations and performing scientific integration, analysis, and information management with software automation tools. Our tool efforts address the problematic issues of very large datasets and varied formats encountered during seismic calibration research. New information management and analysis tools have resulted in demonstrated gains in efficiency for producing scientific data products and improved accuracy of derived seismic calibrations.

The foundation of a robust, efficient data development and processing environment is composed of many components built upon engineered versatile libraries. We incorporate proven industry best practices throughout our code and apply source code and bug tracking management, as well as automatic generation and execution of unit tests, for our experimental, development, and production lines. Significant software engineering and development efforts have produced an object-oriented framework that provides database-centric coordination between scientific tools, users, and data. Over a half billion parameters, signals, measurements, and metadata entries are all stored in a relational database accessed by an extensive object-oriented multi-technology software framework that includes stored procedures, real-time transactional database triggers and constraints, and coupled Java and C++ software libraries to handle the information interchange and validation requirements. Significant resources were applied to schema design to enable management of processing methods and station parameters, responses and metadata. This approach allowed for the development of merged ground-truth (GT) datasets compiled by the NNSA labs and the Air Force Technical Applications Center (AFTAC) that include hundreds of thousands of events and tens of millions of arrivals. The schema design groundwork facilitated extensive quality control and revalidation steps. In support of the GT merge effort, a comprehensive site merge process was also accomplished this year that included station site information for tens of thousands of entries from NNSA labs, AFTAC, National Earthquake Information Center (NEIC), International Seismological Centre (ISC), and Incorporated Research Institutions for Seismology (IRIS). A core capability is the ability to rapidly select and present subsets of related signals and measurements to the researchers for analysis and distillation both visually (JAVA GUI client applications) and in batch mode (instantiation of multi-threaded applications on clusters of processors). Regional Body-Wave Amplitude Processor (RBAP) Version 2 is one such example. Over the past year RBAP was significantly improved in capability and performance. A new role-based security model now allows fine-grain access control over all aspects of the tool's functions enabling researchers to share their work with others without fear of unintended parameter alterations. A new, faster, and more-reliable geographic information system (GIS) mapping framework was added, as well as expanded powerful interactive plotting graphics. In addition, we implemented parent-child type projects to enhance calibration data management.

Our specific automation methodology and tools improve researchers' ability to assemble quality-controlled research products for delivery into the NNSA Knowledge Base (KB). The software and scientific automation tasks provide the robust foundation upon which synergistic and efficient development of Ground-Based Nuclear Explosion Monitoring Research and Development (GNEMRD) Program seismic calibration research may be built.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **SEP 2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE **Enhancing Seismic Calibration Research Through Software Automation and Scientific Information Management**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Lawrence Livermore National Laboratory,PO Box 808,Livermore,CA,94551-0808**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

Proceedings of the 30th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies, 23-25 Sep 2008, Portsmouth, VA sponsored by the National Nuclear Security Administration (NNSA) and the Air Force Research Laboratory (AFRL)

14. ABSTRACT

see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**10**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

OBJECTIVE

The NNSA GNEMRD Program has made significant progress enhancing the process of deriving seismic calibrations and performing scientific integration with automation tools. We present an overview of our software automation efforts and framework to address the problematic issues of improving the workflow and processing pipeline for seismic calibration products, including the design and use of state-of-the-art interfaces and database centric collaborative infrastructures. These tools must be robust and intuitive and reduce errors in the research process. This scientific automation engineering and research will provide the robust hardware, software, and data infrastructure foundation for synergistic GNEMRD Program calibration efforts. The current task of constructing many seismic calibration products is labor intensive, complex, expensive, and error prone. The volume of data as well as calibration research requirements has increased by several orders of magnitude over the past decade. The increase in the quantity of data available for seismic research over the last two years has created new problems in seismic research: data quality issues are hard to track, given the vast quantities of data, and quality information is readily lost if not properly tracked in a manner that supports collaborative research. We have succeeded in automating many of the collection, parsing, reconciliation, and extraction tasks individually. Several software automation tools have also been produced and have resulted in demonstrated gains in efficiency for producing derived scientific data products. In order to fully exploit voluminous real-time data sources and support new requirements for time-critical modeling, simulation, and analysis, continued expanded efforts to provide a scalable and extensible computational framework will be required.

RESEARCH ACCOMPLISHED

The primary objective of the Scientific Automation Software Framework (SASF) effort is to facilitate the development of information products for the GNEMRD regionalization program. The SASF provides efficient access to, and organization of, large volumes of raw and derived parameters, while also providing the framework to store, organize, integrate, and disseminate derived information products for delivery into the NNSA KB.

These next generation information management and scientific automation tools are used together within specific seismic calibration processes to support production of tuning parameters for the United States Atomic Energy Detection System (USAEDS) run by the Air Force. The automation tools create synergy and synthesis between complex modeling processes and very large datasets by leveraging a scalable and extensible database centric framework. The requirements of handling large datasets in diverse formats, and facilitating interaction and data exchange between tools supporting different calibration technologies, has led to an extensive scientific automation software engineering effort to develop an object oriented database-centric framework using proven research-driven workflows and excellent graphics technologies as a unifying foundation.

The current framework supports integration, synthesis, and validation of the various different information types and formats required by each of the seismic calibration technologies. For example, the seismic location technology requires parameter data (site locations, bulletins) and time-series data (waveforms) and produces parameter measurements in the form of arrivals, gridded geospatially registered correction surfaces, and uncertainty surfaces. Our automation efforts have been largely focused on research support tools, RBAP and Knowledge-Base Automated Location Assessment and Prioritization (KBALAP). Furthermore, increased data availability and research requirements have driven the need for multiple researchers to work together on a broad area, asynchronously.

Database-Centric Coordination Framework

As part of our effort to improve our efficiency we have realized the need to allow researchers to easily share their results with one another. For example, as the location group produces GT information, that information should become available for other researchers to use. Similarly, phase arrival picks made by any qualified user should also become immediately available for others to use. This concept extends to the sharing of information about data quality. It should not be necessary for multiple researchers to have to repeatedly reject the same bad data, or worse, miss rejecting bad data. Rather, once data are rejected because of quality reasons they should automatically be excluded from processing by all tools. We are implementing this system behavior using database tables, triggers, stored procedures, and application logic. Although we are at the beginning of this implementation, we have made significant progress over the last year with several kinds of information sharing using the new database centric coordination framework. These are discussed below.

Significant software engineering and development efforts have been successfully applied to construction of an object-oriented database framework that provides database-centric coordination between scientific tools, users, and data.

A core capability of this new framework provides is information exchange and management between different specific calibration technologies, and their associated automation tools, such as seismic location (e.g., KBALAP), seismic identification (e.g., RBAP), and data acquisition and validation (e.g., KBITS). A relational database (Oracle) provides the framework for organizing parameters key to the calibration process from both Tier 1 (raw parameters such as waveforms, station metadata, bulletins, etc.) and Tier 2 products (e.g., derived measurements such as GT, amplitude measurements, calibration, and uncertainty surfaces). Seismic calibration technologies (location, identification, etc.) are connected to parameters stored in the relational database by an extensive object-oriented multi-technology software framework that includes elements of schema design, PL/SQL, real-time transactional database triggers and constraints, as well as coupled Java and C++ software libraries to handle the information interchange and validation requirements. This software framework provides the foundation upon which current and future seismic calibration tools may be based. Interim results and a complete set of working parameters must be available to all research teams throughout the entire processing pipeline. Finally, our development staff has continually and efficiently leveraged our Java code library, achieving 45% code reuse (in lines of code) throughout several thousand Java classes. Source code control is managed by CVS (source code) and ER Studio (schema designs).

Process Improvement

Given the small size of our development staff, the ambitions of our researchers, and the heritage of many of our projects, our process has always been minimal. However, as the complexity and number of users of our system increased, the need for more discipline became apparent. For several years we have been using version control for our source code and have employed unit tests for selected high-risk modules. We have also maintained models of our database artifacts and briefly experimented with maintaining models of our code base. These experiments demonstrated that in our environment of rapidly changing requirements, modeling of source code, except on an as-needed basis, is not practical for us. Because the database does not evolve as rapidly as the code, it is tractable to maintain models of the database, and these prove to be useful both in the design of code and in the refactoring and extension of our database objects.

As the number of users of our applications has increased, it has become increasingly apparent that we must manage multiple system deployments, each with its own application servers and database. There is simply too much content on our production system to risk damage from code under development. Also, disruption to end users is a significant problem when developing on the production system. Accordingly, this year we established two additional CVS branches. One is for experimental development work, and the other is a release branch where we can do bug fixes on our deployed code without interference to ongoing code development. Each branch contains not only the source code, but all artifacts including data models and IDE configurations.

During the development of RBAP 2 we started tracking errors and features on a spreadsheet and discovered, among other things, that we were getting a significant number of regressions in the user interface code. Clearly, just maintaining unit tests for high-risk code wasn’t enough. However, writing a comprehensive set of unit tests would take an amount of effort far greater than what we were devoting to fix regressions. Our compromise solution is to evaluate the utility of automated unit testing. Accordingly, we will soon install the JTest software product from Parasoft that helps automate the generation and execution of a comprehensive suite of unit tests and static analysis.

Another lesson we learned during the development of RBAP 2 was the usefulness of static analysis in eliminating actual or potential errors in the source code. By using automated tools we were able to detect and fix hundreds of problematic or erroneous code constructs. As a consequence, we have agreed on a process by which all code submitted to CVS by developers undergoes a set of static analyses and checks on formatting. When JTest is installed, the static analysis and unit tests will become mandatory upon check-in.

These changes are part of our effort to evolve a software development process with the right balance of agility and discipline for the environment in which we develop and in which our software is used.

Automating Tier 1

Corrections and parameters distilled from the calibration database provide needed contributions to the NNSA KB for the ME/NA/WE region and will improve capabilities for underground nuclear explosion monitoring. The contributions support critical functions in detection, location, feature extraction, discrimination, and analyst review. Within the major process categories (data acquisition, reconciliation and integration, calibration research, product distillation) are many labor intensive and complex steps. The previous bottleneck in the calibration process was in the reconciliation and integration step. This bottleneck became acute in 1998 and the KBITS suite of automated parsing, reconciliation, and integration tools for both waveforms and bulletins (ORLOADER, DDLOAD, UpdateMrg) were developed. The KBITS suite provided the additional capability required to integrate data from many data sources and external collaborations. Data volumes grew from 11,400 events with 1 million waveforms in 1998 to over half a billion raw parameters, measurements, and associated 100 terabytes of continuous data today (e.g., Ruppert et al., 1999; Elliott et al., 2006).

We receive enormous amounts of seismic data daily that must be properly processed. Previously, the movement and management of data were performed manually by our information technology (IT) staff and were extremely time intensive and inefficient. In response, we designed and implemented a distributed (multimachine), multiprocess solution to help automate the collection, movement, cataloging, reporting, viewing, and error processing of waveform segmentation data from multiple academic and government sources. The distributed processes are being written in Java, using encrypted data transfers, logging, an embedded Java relational database (Derby) for maintaining transfer metadata, and a monitoring interface for reporting and quality control. Also, the ability to easily query and view available continuous data was added to improve the efficiency of quality control and recording of metadata.

Automating Tier 2

As the data sources required for calibration have increased in number and source location, it has become clear that the manual, labor-intensive process of humans transferring thousands of files and unmanageable metadata cannot keep the KBITS software fed with data to integrate, nor can the seismic researcher efficiently and consistently find, retrieve, validate, or analyze the raw parameters necessary to effectively produce seismic calibrations in an efficient manner. Significant software engineering and development efforts were applied to address this critical need to produce software aids for the seismic researcher. The principal focus of these efforts has been on the development of two scientific automation tools, RBAP and KBALAP, for seismic location and seismic identification calibration tasks, respectively.

RBAP Version 2

The RBAP (Ruppert et al., 2007) has been extensively revised over the past year. Changes include:

- **Support for tiered projects:** Users can now create a parent calibration project, which defines the velocity model and MDAC parameters for a specific station and region. Subsequent child projects can then be created for the same station and region and MDAC corrected amplitudes calculated using the parent's calibrated parameters. This allows us to explore different source types in understanding discrimination.
- **An improved security model:** Now a minimum-privileged user is able to examine all aspects of any project without being able to change any data. Also, a project owner can now enlist multiple project group members who are able to add measurements without the risk of changing project parameters.
- **Improved scalability:** Computations have been reworked to use multiple producer-consumer threads with the number of threads scaling by available cores.
- **GIS overhaul:** The old GIS system has been replaced by a much higher-performance and more interactive system. Users may now drill-down from the map to view metadata behind measurements.
- **User interface improvements:** The user interface has been reworked to have a consistent look and feel throughout and the level of encapsulation in the GUI code has been substantially increased.
- **Discrimination capability:** Discrimination plots may now be easily produced for any station and any combination of phases and bands.
- **GT editing:** Users with appropriate database privileges may now view and manipulate GT etype information on a per-event basis.
- **Code Quality:** Thousands of code refactorings were performed to better meet industry best-practice standards and patterns.

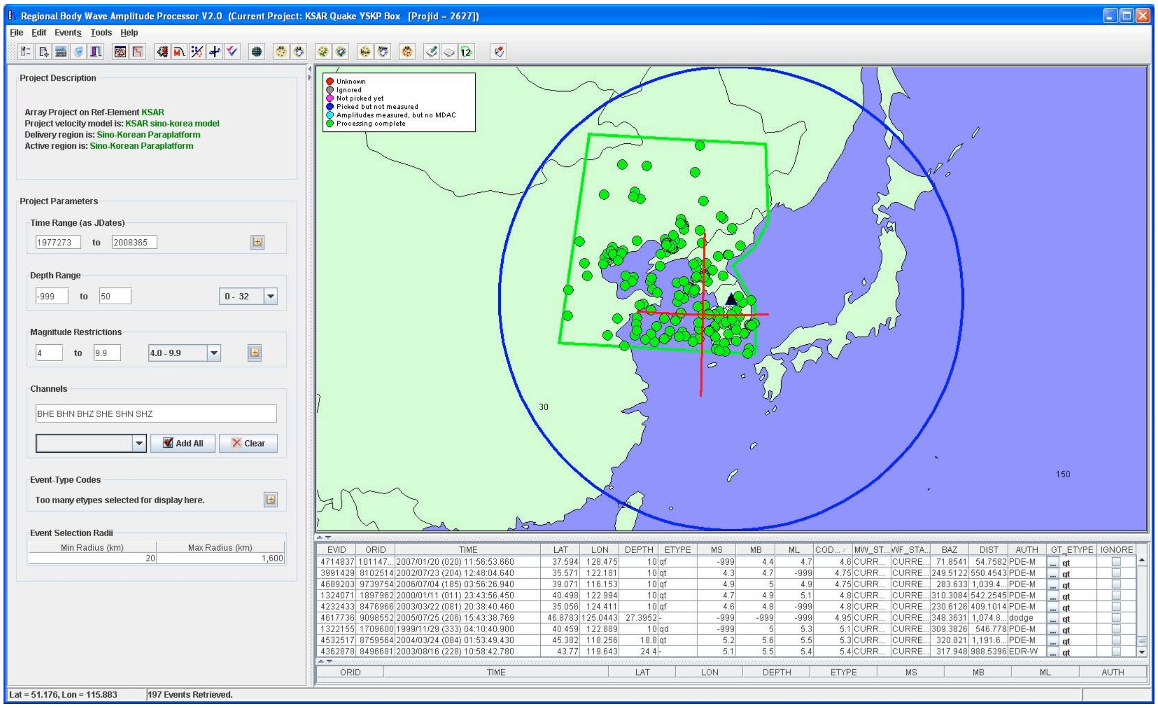

Figure 1. A screenshot of the RBAP program, showing the project parameters in the left pane and a set of processed events displayed in the map and table.

Adding MDAC and Coda Magnitude Processing Module to RBAP

To supplement RBAP in source identification we are developing a waveform processing and signal analysis tool (WFT), which includes the ability to measure amplitudes and coda magnitudes. WFT was originally part of the Hydroacoustic Blockage Assessment Tool (HABAT) code (Matzel et al., 2005) used to discriminate explosion sources from earthquakes in the oceans. WFT has since been broken off into a standalone program with the ability to read, write, plot, and process seismic analysis code (SAC) and CSS format data from flat files or from the LLNL database and was used in the seismic inversion of 3D structure along the Tethyan margin (Flanagan et al., 2006). Along with the filtering, plotting and signal analysis routines derived from the original SAC algorithms, the WFT now includes two subprograms used in seismic discrimination studies: the Amplitude Measurement Tool (AMT) calculates spectral amplitudes, and the Coda Tool calculates coda magnitudes for calibrated regions. The AMT (Figure 2) makes raw and MDAC corrected amplitude measurements on flat file SAC waveforms. This code was designed specifically for use when the LLNL database is inaccessible, and allows us to work more easily with offsite collaborators. AMT performs all the basic RBAP amplitude calculations using the same MDAC parameter setup, it writes output to DiscrimData tables and can plot the results. The Coda Tool (Figure 3) allows anyone with a basic background in coda theory to read either database or flat file seismic data and calculate source magnitudes, given a regional calibration. Once data are read in, the user can calculate seismic data envelopes, calculate synthetics based on published theory (Mayeda et al., 2003), compute the spectral amplitudes, add site and path corrections, and compute the final coda Mw. We are currently extending the Coda Tool to create regional calibrations and to work directly with the LLNL database.

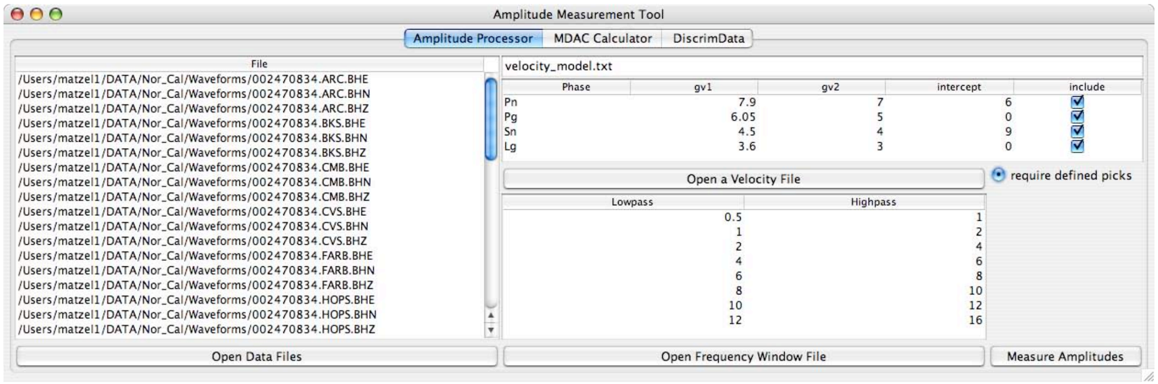

Figure 2. A screenshot of the Amplitude Measurement Tool, which calculates spectral amplitudes for the seismic analysis code (SAC) and CSS flat files and applies MDAC corrections to the results.

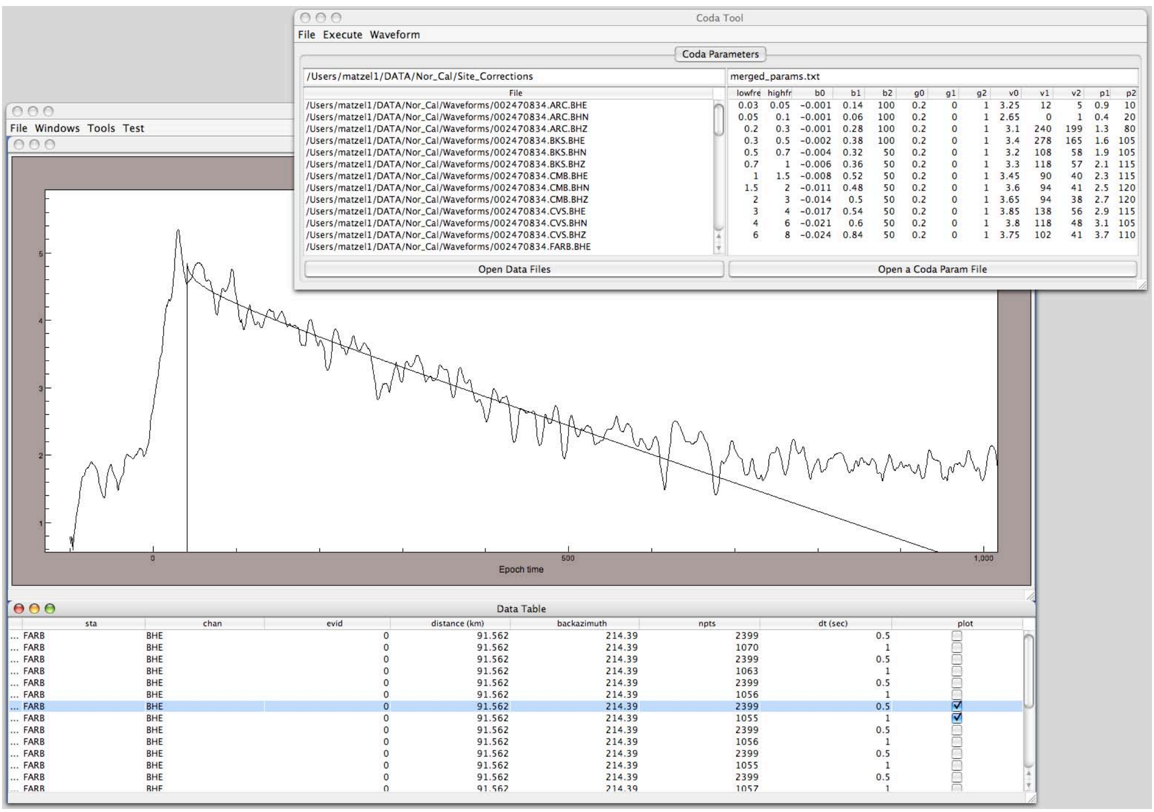

Figure 3. A screenshot of the Coda Mw measurement tool, illustrating the measured data envelopes compared with synthetic envelopes.

The KBALAP Program

The KBALAP program is another Tier 2, event-centric automation effort in the GNEMRD program (Elliott et al., 2006). It is a highly interactive, graphical tool that uses a set of database services and a client application based on data selection profiles that combine to efficiently produce location GT. These data can be used in the production of travel time correction surfaces, and as part of the preferred event parameters used by other tools in our processing framework.

KBALAP's database services are responsible for evaluating bulletin and pick information as it enters the system for identifying origin solutions that meet predefined GT criteria without further processing and for identifying events that would likely meet a predefined GT level if a new origin solution were produced using available arrivals. The database service is also responsible for identifying events that should have a high priority for picking based on their existing arrival distribution, and the availability of waveform data for stations at critical azimuths and distances.

The interactive portion of KBALAP has the following principal functions:
• Production of GT origins through prioritized picking and location
• Specification of GT-levels for epicenter, depth, origin time and etype
• Batch-mode location of externally produced GT information
• Production of array azimuth-slowness calibration data
• Easy review and modification of event parameters used by all GNEMRD researchers

Some key KBALAP features are listed below:
• Fast and efficient location
• Project management and collaboration
• Batch processing

The Site Merge Effort

Information about seismic station position and installed instrumentation is to a greater or lesser extent, fundamental to all the processing done within the GNEMRD Program. However, despite the importance of accurate information about seismic stations, in practice it is difficult to obtain a compilation of station information that does not include errors. There are many sources for these errors, including the following:
• Imprecise surveying/reporting by station operators
• Transcription errors
• Unrecorded station movements or equipment modifications

The situation is complicated even more by the fact that many different compilations have been produced using different sources and different assumptions, and these compilations are inconsistent with one another.

In the past, we have dealt with inconsistencies case by case. When a problem was identified, we would “fix” the offending data in our SITE table and go on. While this approach was problematic in a number of ways, given the limitations of the CSS SITE table and our need to build out other parts of our infrastructure, it was judged to be the best we could do. As the labs coordinate more in the process of producing calibration products for monitoring purposes, the need for a unified, consistent SITE table has become more apparent. Producing and maintaining such a table by integrating and reconciling our individual SITE tables is an even more difficult undertaking than simply maintaining an internal-use-only SITE table. Mainly this is because of the need to resolve conflicts in a way that is trackable, reproducible, and backed with documented decisions/assumptions.

We were tasked this year with performing the location GT merge between contributing laboratories. This effort depended critically on having a unified SITE table of the highest possible quality. This accelerated our work on producing a SITE merge, and we now have a system that we used to produce an integrated SITE table for the GT merge and as a replacement for our LLNL SITE table. Our merge process is implemented in Java and in PL/SQL and uses a number of tables to track metadata about the merge process. The codes allow for repeated contributions by the same author allowing, for example, updating of the merged SITE as new versions of the NEIC station book become available. The results and documentation will be provided to the relevant NNSA GNEMRD working groups for coordination and consideration.

Our approach to merging SITE data is to handle the position, elevation, operating epochs, station movements, array membership and possible code aliasing separately. We take this approach because there is no guarantee that a particular contributor’s information about a SITE will be uniformly better or worse than information from another source.

When SITE data come into the system, they are placed into a multiauthor site table (and supporting tables) that hold all the unmerged data. Before a new merge is executed, a process is run that identifies unresolved discrepancies (over a threshold value) in position and elevation. Any stations with unresolved discrepancies are added to appropriate discrepancy tables. Although the merge can continue without resolving the discrepancies, these stations will not become part of the merged SITE table.

Discrepancies can be resolved in one of two ways: by making entries in a preferred table or by making entries in a rejected table. The reason column in each of these tables allows up to a 2000-character discussion of the reason for the decision. With this system, it is relatively easy to find out why a particular datum was or was not used, and if better information becomes available it is easy to change the first decision and redo the merge. The software also helps resolve position discrepancies by producing KML files that allow display in Google Earth of clusters of discrepant station position estimates.

We handle alternate station codes by excluding them from the merge process. After the merge is complete, duplicate entries are created for each alternate code, but with the original code replaced by the alternate code. In addition to the alternate codes provided by the NEIC, we identified a number of additional such codes by comparison of latitude, longitude, and elevation entries. These are treated in the same fashion as the NEIC alternate codes.

We have used our SITE merging system to combine SITE information from the most recent NEIC and ISC station books, the current NNSA GNEMRD SITE tables, and the Incorporated Research Institutions for Seismology (IRIS) SITE table (derived from dataless SEED volumes minus temporary deployments and California stations). There are about 39,000 entries in the multiauthor site table which produce more than 17,700 merged SITE entries. There are 376 preferred positions, 286 preferred elevations, 61 rejected positions, and 572 rejected elevations. The position overrides were determined mostly through a combination of inspection in Google Earth and residual analysis using GT events. Most of the elevation overrides were arrived at by comparison of reported elevations with elevations computed using the gtopo30 elevation model.

Ongoing Work Related to the SITE Merge

Although the code for producing the merged SITE is fully functional, we do not yet have an automated means of updating our production database when a new SITE table is produced. The problem is that if stations that exist in the current table are updated and their epochs change, it affects many other tables that have some dependence on SITE. This time we resolved those issues manually, but that process must be automated for future SITE updates. Also, it is apparent that we need to develop software tools that will allow a nontechnical user to make changes to the SITE table. The software must not only make changes to the production table(s), but must make appropriate entries in the metadata schema as well. For example, changing a station position must now require review of any justifications for the current position, and it must be required that the change be accompanied by a justification. After accepting the change, the system must update all tables affected by the change.

Instrument Response Updates

We are in the process of converting most of our SEED response files to the poles and zeros/finite impulse response (PAZFIR) format. The conversion is being done using code developed by George Randall of LANL. Each converted response is checked by comparing frequency-amplitude-phase spectra generated from the PAZFIR data to that generated from the SEED data. Those with an RMS difference of less than 1% (virtually all of the conversions) are retained. The remainder will stay as SEED for now.

Roughly 3,600 of our existing IRIS dataless SEED RESP files have been successfully converted, resulting in over 7,000 PAZFIR responses. In order to properly convert and validate responses, this effort involved writing C-shell, Java, and Matlab programs (that used Perl programs to perform the actual conversion). The principal effort remaining is to modify the conversion codes to move the converted responses into appropriate directories and to update the appropriate instrument rows.

The GT Merge Effort

The GT merge project involved the combination of GT data sets compiled by LLNL and LANL along with supporting and associated data from the labs and from AFTAC. The merge process also included extensive quality-control and revalidation steps. In all, over 230,000 events with over 20,000,000 arrivals were processed.

This process merges the GT25 and better datasets between contributing laboratories for use in both a tomographic inversion for *Pn* velocity of Eurasia and for computing first-*P* correction surfaces using the KBCIT software. The merge is intended both to resolve GT common between labs (choosing the better GT estimate when possible) and to perform an extensive set of quality control steps to the origin and phase data. We have developed a software system implemented in Java, Oracle relational database, and PL/SQL to perform this merge process (Flanagan et al., 2007).

The software brings together into a GTMERGE schema the GT data from both labs along with all supporting ORIGIN, ORIGERR, ASSOC and ARRIVAL data. All data are given new IDs unique within the GTMERGE schema, and events in common between labs are identified by spatial-temporal correlation. The Bondar-Myers-Engdahl-Bergman (BMEB) Epicenter accuracy criteria are used (Bondar et al., 2004). For those events in common, a set of ranking rules is applied to select the best non-BMEB GT. A small subset of the input GT that cannot be ranked is resolved manually.

The quality check (QC) steps performed by the software include the following:
• Enforcing common phase naming conventions
• Removing arrivals that are too early or too late to be of interest
• Removing phases not of interest
• Identifying and removing arrivals that are too discrepant to be useful
• Enforcing distance-dependent phase name conventions
• Choosing a "best" arrival for each EVID-STA-PHASE table

After QC is complete, the system evaluates all the BMEB GT for strict adherence to their criteria. All events that fail this check have a new origin solution computed using phase gathers appropriate to the GT level. If the new solution meets its criterion, then it is included in the final merge results. Otherwise, the event is rejected. When all GT have been re-evaluated, a new set of constrained origin solutions is computed using teleseismic P-arrivals. These "baselined" origins are the final product of the merge effort. The dataset produced by the merge effort includes about 97,000 distinct events with nearly 20,000,000 arrivals.

CONCLUSIONS AND RECOMMENDATIONS

We present an overview of our software automation efforts and framework to address the problematic issues of consistent handling of the increasing volume of data, collaborative research efforts and researcher efficiency, and overall reduction of potential errors in the research process. By combining research-driven interfaces and workflows with graphics technologies and a database centric information management system coupled with scalable and extensible cluster-based computing, we have begun to leverage a high performance computational framework to provide increased calibration capability. These new software and scientific automation initiatives will directly support our current mission including rapid collection of raw and contextual seismic data used in research, provide efficient interfaces for researchers to measure and analyze data, and provide a framework for research dataset integration. The initiatives will improve time-critical data assimilation and coupled modeling and simulation capabilities necessary to efficiently complete seismic calibration tasks. This GNEMRD Program's scientific automation, engineering, and research will provide the robust hardware, software, and data infrastructure foundation for synergistic calibration efforts.

ACKNOWLEDGEMENTS

We acknowledge the assistance of the LLNL computer support unit in implementing and managing our computational infrastructure. We thank Jennifer Aquilino, Laura Long, and John Hernandez for their assistance in configuration and installation of our Linux cluster and workstations.

REFERENCES

Bondar, I., S. Myers, E. Engdahl, and E. Bergman (2004). Epicenter accuracy based on seismic network criteria, *Geophys. J. Int.* 156: 483–496.

Elliott. A., D. Dodge, M. Ganzberger, T. Hauk, E. Matzel, and S. Ruppert (2006). Enhancing seismic calibration research through software automation and scientific information management, in *Proceedings of the 28th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-06-5471, Vol. 2, pp. 967–975.

Flanagan, M, D. Dodge, and S. Myers (2007). GT merge processing documentation—Beta version, in University of California, Radiation Laboratory, 1–15 (in review).

Flanagan, M., E. Matzel, S. van der Lee, H. Bedle, M. Pasyanos, F. Marone, B. Romanowicz, C. Schmid, and A. Rodgers (2006). Joint inversion for three-dimensional velocity structure of the broader Africa-Eurasia collision region, in *Proceedings of the 28th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-06-5471, Vol. 1, pp. 397–406.

Matzel, E., A. L. Ramirez, and P. E. Harben (2005). Model-based hydroacoustic blockage assessment and development of an explosive source database, in *Proceedings of the 27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies,* LA-UR-05-6407, Vol. 2, pp. 733–739

Mayeda, K., A. Hofstetter, J. L. O'Boyle, and W. Walter (2003). Stable and transportable regional magnitudes based on coda-derived moment-rate spectra, *BSSA* 93: 224–239.

Ruppert, S., T. Hauk, J. O'Boyle, D. Dodge, and M. Moore (1999). Lawrence Livermore National Laboratory's Middle East and North Africa research database, in *Proceedings of the 21st Seismic Research Symposium: Technologies for Monitoring the Comprehensive Nuclear-Test-Ban Treaty*, LA-UR-99-4700, Vol. 1, p. 243.

Ruppert, S., D. Dodge, M. Ganzberger, T. Hauk, and E. Matzel (2007). Enhancing seismic calibration research through software automation and scientific information management, in *Proceedings of the 29th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-07-5613, Vol. 2, pp. 965–974